

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **2000293420 A**

(43) Date of publication of application: **20.10.00**

(51) Int. Cl.

**G06F 12/00**

(21) Application number: **11102167**

(22) Date of filing: **09.04.99**

(71) Applicant: **HITACHI LTD**

(72) Inventor: **KIHARA KENICHI  
HIRATA TOSHIAKI  
SAITO TARO**

(54) **DATA BACKUP SYSTEM**

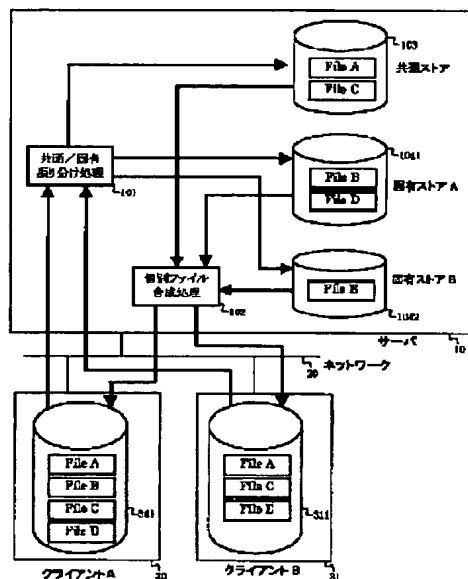
(57) Abstract:

PROBLEM TO BE SOLVED: To decrease the total capacity of backup data by detecting common parts in data (file group) that respective clients have and saving (backing up) the common parts between the data of the clients and characteristic parts of the data of the respective clients separately.

SOLUTION: This system comprises four components which are a server 10, a network 20, and clients A30 and B31. The clients A30 and B31 are connected to each other through the server 10 via the network 20. The clients A30 and B31 have files locally and transmit the respective files to the sever 10 when they are backed up. The server 10 divides the files sent from the clients into common parts and characteristic parts by a common/characteristic part classification process 101 and stores the common parts in a common storage 103, and the characteristic parts of the client A30 in a client-A characteristic part storage

A041 and the characteristic parts of the client B31 in a client-B characteristic storage A042.

COPYRIGHT: (C)2000,JPO



(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開2000-293420

(P2000-293420A)

(43)公開日 平成12年10月20日(2000. 10. 20)

(51)Int.Cl.<sup>7</sup>

G 0 6 F 12/00

識別記号

5 3 1

F I

G 0 6 F 12/00

テーマコード(参考)

5 3 1 M 5 B 0 8 2

5 3 1 R

審査請求 未請求 請求項の数 5 O L (全 21 頁)

(21)出願番号

特願平11-102167

(22)出願日

平成11年4月9日(1999. 4. 9)

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 木原 健一

神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(72)発明者 平田 俊明

神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(74)代理人 100075096

弁理士 作田 康夫

最終頁に続く

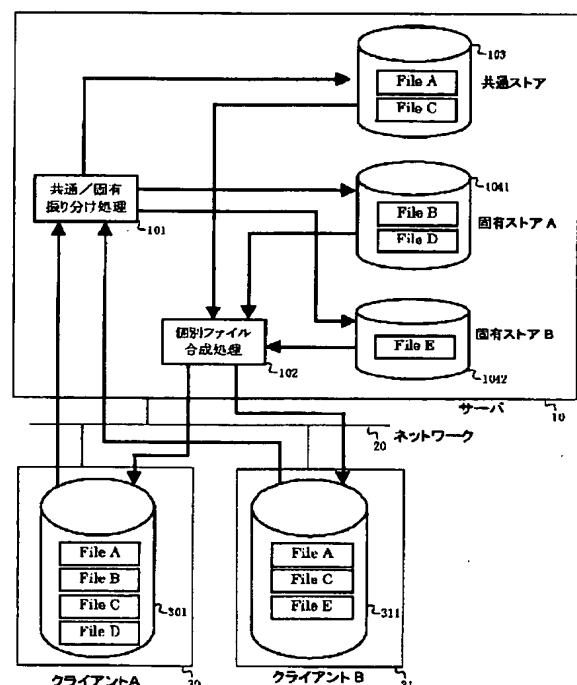
(54)【発明の名称】 データバックアップシステム

(57)【要約】

【課題】クライアントのバックアップファイルをサーバに保存する環境において、クライアントの台数が多数の場合にサーバ上のディスクスペースが大量に必要なになる。

【解決手段】クライアントのバックアップファイルをクライアント間で比較し、同一部分と固有部分とに分けて格納し、サーバのディスクスペースを節約する。

図1.



**【特許請求の範囲】**

**【請求項 1】** 複数のクライアント機器と、サーバ機器とからなるネットワークシステムにおけるデータバックアップシステムであって、前記サーバ機器は、前記各クライアント機器が持つデータから各クライアント機器に共通なデータ部分と各クライアントに固有なデータ部分とを検出する手段と、前記共通なデータ部分を保存する共通データ保存手段と、前記固有なデータ部分を保存する固有データ保存手段とを備えることを特徴とするデータバックアップシステム。

**【請求項 2】** 請求項 1 記載のデータバックアップシステムにおいて、前記共通データ保存手段に保存されたデータそれぞれが、どのクライアント機器のデータであるかを示すデータを記憶する手段を備えることを特徴としたデータバックアップシステム。

**【請求項 3】** 請求項 1 記載のデータバックアップシステムにおいて、前記蓄積されたデータを、前記クライアント機器別にリストアする手段を備えることを特徴としたデータバックアップシステム。

**【請求項 4】** 請求項 1 記載のデータバックアップシステムにおいて、前記クライアント機器は、バックアップ対象となるデータの属性情報を前記サーバ機器へ送信する手段を備えることを特徴としたデータバックアップシステム。

**【請求項 5】** 請求項 4 記載のデータバックアップシステムにおいて、前記サーバ機器は、各クライアント機器から送信された前記データ属性情報にもとづいて、データバックアップ時に前記クライアント機器から前記サーバ機器に転送する必要があるファイルを決定する手段を備えることを特徴としたデータバックアップシステム。

**【発明の詳細な説明】****【0001】**

**【発明の属する技術分野】** 本発明はネットワークシステムにおける複数のクライアント機のバックアップ（データの保存）・リストア（データの復旧）方法に関する。

**【0002】**

**【従来の技術】** ネットワーク／システム管理の分野では TCO（総保有コスト）の低減が課題になっており、クライアント機の管理コストが TCO 低減の足かせになっている。特にディスククラッシュからの復旧には、OS やアプリケーションソフトの再インストールや環境設定等、膨大な手間がかかり、管理コスト上昇の元凶となっている。

**【0003】** 従来からディスククラッシュからの復旧に対応するためには、バックアップが有効であることは知られているが、ネットワークシステムにおけるバックアップについても例えば特開平2-297643号公報に記載され

ているような技術がある。この技術は複数のクライアント（ワークステーション）機のバックアップを目的としたもので、バックアップ用のサーバを設置し、クライアント上のバックアップファイルをバックアップ用サーバに蓄積する。

**【0004】**

**【発明が解決しようとする課題】** 前記技術によると、サーバ上でのデータバックアップ領域のサイズは、全クライアント機（ワークステーション）のバックアップデータの総和となり、クライアント台数が多いとサーバ上のデータバックアップ領域も巨大なものが必要となる。バックアップに要するディスク容量を考え、サーバのみバックアップを行いクライアントのバックアップをあきらめてしまうと、結果的にディスククラッシュから復旧できないという問題が生じることになる。

**【0005】** 本発明の目的は、複数のクライアント機が使用される環境において、クライアント機の記憶装置のバックアップデータサイズを縮小することであり、さらにバックアップへの障害をなくすことでバックアップを促しディスククラッシュからの復旧に対応できるようにすることであり、さらにシステムの管理コストを低減することである。

**【0006】**

**【課題を解決するための手段】** 上記目的を達成するために本発明では、各クライアントが持つデータ（ファイル群）内の共通部分を検出し、クライアントのデータ間に共通な部分と各クライアントのデータに固有な部分に分けて保存（バックアップ）することで、バックアップデータの総容量を低減する。

**【0007】** 同じネットワークシステム内にあるクライアント機それぞれが持つファイルまたはファイル群は、互いに同じような構成、内容になることが多い。その結果、各クライアントのバックアップデータの内容も似通った内容になりやすい。本発明はその特性を利用する。

**【0008】** すなわち本発明は、記憶手段を持つ機器における一つあるいは複数のデータで構成されるバックアップデータを蓄積する手段において、異なる前記機器用の前記バックアップデータに含まれるデータの内容を比較し、前記データ内容に同一のものが発見されれば、前記内容が同一のデータを一つのデータに集約して蓄積する事の特徴とするものである。

**【0009】** さらに本発明は、前記機器用の前記同一の単数または複数のデータを蓄積する記憶手段と、前記同一のデータ以外の前記機器用の単数または複数のデータを蓄積する記憶手段を持つことを特徴とするものである。

**【0010】** さらに本発明は、前記機器用の単数または複数のデータを蓄積する手段すなわち共通データ蓄積手段と、前記機器別に前記共通データ蓄積手段に含まれるデータの中のどのデータが前記機器用バックアップデー

タに含まれているかを示す情報を蓄積する記憶手段すなわちデータ所有情報記憶手段とを、持つことを特徴とするものである。

【0011】さらに本発明は、前記バックアップデータを蓄積する手段に蓄積された一つあるいは複数のデータを、前記機器別に取り出す機能を持つことを特徴とするものである。

【0012】さらに本発明は、前記バックアップデータを蓄積する手段すなわちサーバと、データバックアップの対象である前記機器であり、前記サーバとの間で情報をやりとりする機能を持つ手段すなわちクライアントを持つことを特徴とするものである。

【0013】さらに本発明は、前記クライアントと前記サーバの間でデータ名やデータサイズ、データの更新日時等のデータ属性情報を送信することを特徴とするものである。

【0014】さらに本発明は、前記クライアントと前記サーバ間でやりとりされた前記データ属性情報にもとづいて前記クライアントから前記サーバに転送する必要のあるデータを決定することを特徴とするものである。

【0015】さらに本発明は、前記クライアントと前記サーバの間で巡回冗長符号やチェックサム等データ内容から計算されるチェックコードをやり取りすることを特徴とするものである。

【0016】さらに本発明は、前記クライアントと前記サーバ間でやりとりされた前記チェックコードにもとづいて前記クライアントから前記サーバに転送する必要のあるデータを決定することを特徴とするものである。

【0017】さらに本発明は、前記クライアントから前記サーバに情報を伝達する手段とは異なる手段で前記サーバから前記クライアントに情報を伝達することを特徴とするものである。

【0018】さらに本発明は、前記サーバから前記クライアントへの情報伝達手段として、記憶媒体を用いることを特徴とするものである。

【0019】さらに本発明は、前記サーバから前記クライアントへの情報伝達手段として、通信衛星等の大容量のデータ転送に向く通信手段を用いることを特徴とするものである。

#### 【0020】

【発明の実施の形態】 1. 第一の実施例：基本構成

図1はその原理を示した図である。サーバ10、ネットワーク20、クライアントA30、クライアントB31の四つの構成要素からなる。図1では、クライアントは二台表記しているが、一台あるいは三台以上でも問題ない。クライアントA30、クライアントB31はサーバ10とネットワーク20経由で接続されている。

【0021】各クライアントはローカルにファイルを持っており、バックアップの際にそれぞれそのファイルをサーバ10に送信する。サーバ10は各クライアントから送

られたファイルを共通／固有振り分け処理101で、共通部分と固有部分とに分け、共通部分を共通ストア103に、クライアントA30の固有部分はクライアントA固有ストアA041、クライアントB31の固有部分はクライアントB固有ストアB042に保存する。

【0022】たとえば図1の実施例では、クライアントA30はファイル装置301内にFile A、File B、File C、File Dという四つのファイルを持っていて、クライアントB31はFile A、File C、File Eという三つのファイルを持っている。両クライアントに共通するFile AとFile Cについては、共通ストア103に格納される。また、クライアントA30が持つファイル装置301内のFile B、File DについてはクライアントA30固有なので、固有ストアA041に格納する。同様にクライアントB31のファイル装置311内のもFile Eも固有ストアB042に格納する。File AとFile Cについては、別々に格納するのではなく、一つにまとめることで保存領域を節約できる。

【0023】バックアップされたファイルを元のクライアントに戻す場合には、共通ストア103内のファイルとクライアント固有ストアの内容を合わせることで元のファイル（群）が得られる。そのファイル合成を行うのが個別ファイル合成処理102である。たとえばクライアントA30のファイル装置301のリストアが必要になったら、共通ストア103内のFile A、File CとクライアントA固有ストアA041内のFile B、File DをクライアントA30用のリストアデータとする。

【0024】図2はネットワークシステム内で使われるサーバ装置およびクライアント装置のハードウェア構成の例である。内部はCPU41、メモリ42、キーボード43、ディスプレイ44、ハードディスクドライブ45、ネットワークインタフェース46の部品で構成されている。ネットワークインタフェース46は、他のマシンとネットワーク20で接続されている。

【0025】図3は内部処理方法を説明する図で、図1のデータの流れを示したものである。クライアントA30はファイル装置301、ファイル送信機能302、ファイル受信機能303を持つ。ファイル装置301はクライアントA30で使われるファイルを保持するもので、前記ハードディスクドライブ45に相当する。ファイル送信機能302はファイルのバックアップを行う際に、サーバ10にファイル装置301内のファイルを送信する機能を持つもので、前記メモリ42上に前記CPU41が実行するプログラムとして実現される。その逆にファイル受信機能303はファイルのリストアを行う際にサーバ10からファイルを受け取り、ファイル装置301に格納する機能を持つもので、前記メモリ42上に前記CPU41が実行するプログラムとして実現される。

【0026】サーバ10はクライアントA30から送られてくるファイルを、共通／固有振り分け処理101（前記CPU41が前記メモリ42上で実行するプログラムとして実

現される)内のファイル受信1011で受け取り、クライアント別にクライアントA用受信バッファ1012またはクライアントB用受信バッファ1013に格納する(受信バッファはバックアップが必要となるクライアントの台数分を前記ハードディスクドライブ45に用意すればよい)。各受信バッファに格納されたファイルはAND1014で共通部分を抽出され、前記ハードディスクドライブ45内の共通ストア103に格納する。また、SUB1015によってクライアントA用受信バッファ1012から共通部分を差し引いたものを前記ハードディスクドライブ45内の固有ストアA1041に格納する。同様にSUB1016によってクライアントB用受信バッファ1012から共通部分を差し引いたものを前記ハードディスクドライブ45内の固有ストアB1042に格納する。

【0027】またリストアの際には、個別ファイル合成処理102(前記CPU41が前記メモリ42上で実行するプログラムとして実現される)において、リストアが必要なクライアントに応じて共通ストアと固有ストアを合成したものを、ファイル送信1021を用いてクライアントに送信する。たとえばクライアントA30にリストアが必要な場合には、ADD1022を使って共通ストア103と固有ストアA1041の内容を合わせてクライアントA30に送信する。

【0028】図4はクライアントとサーバ間のバックアップのプロトコルを示したものであり、上述の各処理プログラムによって実行される。バックアップの際にはクライアントA30からサーバ10に対してファイルバックアップ情報50を送信する。ファイルバックアップ情報50の中は、コマンドID(バックアップ操作を示す識別コード)、クライアント特定用のIDの後に各ファイルのデータを続ける。各ファイルのデータの内容はファイルパス名、ファイル属性、ファイル本体などを含めればよい。

【0029】図5はクライアントとサーバ間のリストアのプロトコルを示したものであり、上述の各処理プログラムによって実行される。リストアの際には、クライアントA30からリストア要求51をサーバに発行する。リストア要求51の内容にはコマンドID(リストア要求を示す識別コード)とクライアントIDを含める。リストア要求51を受けたサーバ10はリストア要求51内のクライアントIDを元に、前に示した方法でクライアント用ファイルを作成し、リストアファイル情報52として、クライアントに送信する。リストアファイル情報52は、コマンドID(リストアファイル情報であることを示す識別コード)にファイル本体が続く。

【0030】図6はサーバ側のメイン処理の流れを示したものであり、上述の各処理プログラムによって実行される。クライアントからのコマンドを受信し、受信したコマンドが終了指示なら終了し、バックアップ情報50ならバックアップ処理を、リストア要求51ならリストア処理を実行する。

【0031】図7はサーバ側のバックアップ処理の内容を示したものであり、上述の各処理プログラムによって実行される。クライアントから送られてきたファイルをクライアント受信バッファにスプールする。前述の通り、クライアントA用受信バッファ1012またはクライアントB用受信バッファ1013に格納する。次に、すべての受信バッファの内容を確認し、全クライアントからのバックアップファイルが集まったかどうかを判断する。全部集まっていなければ終了する。集まっていれば全受信バッファに共通するファイルを抽出し、共通ストア103に格納する。また、各受信バッファ固有のファイルは固有ストアA1041と固有ストアB1042に格納する。

【0032】図8はサーバ側のリストア処理の流れを示したものであり、上述の各処理プログラムによって実行される。共通分のファイルとリストア要求を出したクライアントの固有ストアのファイルとをリストア要求を出したクライアントへ送信する。サーバ10内でのファイルの共有状況は内部処理なので直接目に見えない。そこで共有状況を画面に表示すれば共有の効果を目視で確認できるようになる。

【0033】図9は共有の効果を確認するため図2のディスプレイ44上に、共通/固有振り分け処理101によって表示される画面例で、固有格納分と共有格納分のファイルサイズを表示している。オリジナル総ファイルサイズ901はクライアント側のファイルの合計(共有なし)、固有格納分はクライアント固有ストア(図3では1041,1042)の合計サイズ、共有格納分は共通ストア103のサイズをそれぞれ表示している。また削減率904は固有格納分902と共通格納分903を足したものをオリジナル総ファイルサイズ901で割った値を表示しており、ファイル共有による効果の度合いを示している。図10は図2のディスプレイ44上に、共通/固有振り分け処理101によって表示されるクライアント別共有比率の表示画面例で、クライアント別906に何パーセントのファイルが共有されているかを、ファイルの個数907とファイルのサイズ908の観点で表示している。図11は図2のディスプレイ44上に、共通/固有振り分け処理101によって表示される共有状況の表示画面例で、共通ストア内のファイルおよび固有ストア内のファイルをそれぞれ表示することで、どのファイルが共有されているかを目視で確認できる。

【0034】上述のように、第一実施例によれば、クライアント機のバックアップデータをサーバに集め、バックアップデータをクライアント間で比較し、同じファイルがあれば一つに集約して保存することで、サーバ上のディスクスペースを節約できる。

【0035】2. 第二の実施例

第二の実施例として、サーバ10上でのバックアップ用ファイルの格納方法を変更した案を図12に示す。第一の実施例(図3)の共通/固有振り分け処理101、個別ファ

イル合成処理102、クライアントA固有ストアA041、クライアントB固有ストアA042に代わり、第二の実施例ではマルチクライアントファイル格納処理1051（前記CPU41が前記メモリ42上で実行するプログラムとして実現される）、オーナー対応表1052（前記メモリ42上に実現される）、マルチクライアントファイル取り出し処理1053（前記CPU41が前記メモリ42上で実行するプログラムとして実現される）を備えている。なお第二の実施例はサーバ10のみの変更で、クライアントA30には変更は必要ない。

【0036】オーナー対応表1052は、共通ファイル103内のファイルについてどのクライアントがそのファイルを所有しているかを示すデータで図13に例示するように、共通ストア上のファイル毎に、ファイルID10521、オーナー10522、ファイルパス名10523が記されており、オーナー対応表1052内の横一行が一レコードになっている。

【0037】ファイルID10521はオーナー対応表1052と共通ストア103の結びつけを行うためのもので、そのレコードが共通ストア103上のどのファイルに対するものかを示すための情報となっている。図13の例では、ファイルID10521は1001や1002などの数値で示しているが、区別可能なら記号でもよい。オーナー10522はそのファイルがどのクライアントによって所有されているかを示している。ファイルパス名10523はそのファイルがクライアントのファイル装置301上でどこ（のファイルパス）に配置されていたか（および配置させるか）を示すものである。

【0038】図14はオーナー対応表1052の他の例である。図13では共通ストア103内の一ファイルに対してオーナー対応表1052の一行（一レコード）が対応していたが、図14では共通ストア103内の一ファイルがオーナー対応表1052の複数行に対応する。こうすることで、ファイルパス名10523（あるいはタイムスタンプ10524もしくは属性10525）は異なるがファイルの内容は同じ場合に、図13の例では複数のファイルが共通ストア103に格納されることになるが、図14の例では一つで済む。図14の例では、ファイルIDが1001のファイルは、クライアントAのファイル装置103上ではC:\DIR1\FI  
LE01の場所に置かれているが、クライアントBではD:\DI  
R2\FI  
LE01の場所に置かれていることを示している。

【0039】マルチクライアントの格納処理1051の内容を図15に例示する。クライアントからファイルを受信すると、オーナー対応表1052からその送信元のクライアントの情報を取り除く。たとえばクライアントBからファイルを受信したとき、図14の例の場合、（図14に現れている範囲内で）二行目と三行目のレコードが削除される。レコードを削除した結果、オーナーが存在しなくなった共通ストア103内のファイルは、共通ストア103から削除する。以下、クライアントから受信した各ファ

イルについて次の処理を行う。すなわち、すでに同じファイルが共通ストア103上に存在するかどうかを判断し、共通ストア103内になければそのファイルを共通ストア103に格納し、共通ストア103内にある・ないに関わらずオーナー対応表1052の内容を更新する。

【0040】第一の実施例と同様に、サーバ10内での共有状況を表示する仕組みを持たせると利用者が共有状況を確認できるので良い。図16はその画面例である。統計情報911として、オリジナル総ファイルサイズ（クライアントのファイルのサイズの合計）、格納ファイルサイズ（共通ストア103内に格納されているファイルのサイズの合計）、節約できたバイト数（前者二つの差）を示している。また共通ストア103内に格納されているファイルの一覧912として、共通ストア103内の各ファイルについてそれぞれファイルサイズとオーナー数を表示している。

【0041】この第二の実施例によれば、あるファイルが総てのクライアントに共通でなくても二台以上に共通していれば共有可能になる。そのため、第二の実施例の方がよりディスクスペースを節約できる可能性がある。

【0042】第二の実施例において、クライアントから送られてきたファイルと同じ内容のファイルが、すでに共通ストア103内にあるかどうかを判定する方法について説明する。クライアントから送られてきたファイルのそれぞれについて、共通ストア103内に格納されているファイル（群）の先頭から順に比較を行う手法を使った場合、次の二つの要因で処理量が膨大になる。

【0043】要因1．比較回数が膨大  
共通ストア103内の先頭から順に比較を行うので、クライアント側のファイル一つあたり、最悪で共通ストア103内のファイル個数回の比較が必要になる。平均でもファイル個数の半分程度の比較回数が必要になる。クライアント側の全ファイルに対してその処理を行う場合、平均で、「クライアント側の全ファイル」\*「共通ストア内のファイル個数」/2回の比較が必要になる。たとえば、クライアントから1,000個のファイルが送られて、共通ストア103内に10,000個のファイルが存在する場合、比較回数は、最悪10,000,000回程度の比較が必要になる。

【0044】要因2．比較一回あたりの処理量が膨大  
ファイル内容同士を比較するため、比較一回あたりの処理量も大きくなる。

【0045】そこで、比較回数を低減する方式について説明する。

【0046】比較一回あたりの処理量が大きい点については、ファイル本体を比較する前にファイルパス名やファイルの長さ、チェックサム等を比較することで処理量を低減できる。また、比較回数が多い点は、オーナー対応表や共通ストア103内の情報をファイル名等でソートしておくことで、二分探索法などの効率の良い検索（比

較)が行えるようになる。

【0047】ファイルの長さをキーとしたファイル比較の効率化手法の例を一つ挙げる。本効率化手法例ではファイル探索を効率化するために、ファイル早見表109を用意する。ファイル早見表109の構造を図17に示す。ファイル早見表109は単数あるいは複数のファイル早見情報1091で構成されており、各ファイル早見情報1091はさらにファイルサイズ10911、チェックサム10912、ファイルID10913を含む。ファイル早見情報1091の並びは、ファイルサイズを主キー、チェックサムを副キーとしてソートされている。すなわち基本的にファイルサイズでソート済みで、ファイルサイズが同じファイル早見情報1091同士はチェックサムでソートする。クライアントから送られてきた個々のファイルのファイルサイズとチェックサムを比較対象の絞り込みに使用する。

【0048】図18はファイル早見表109を用いた探索手法の例である。この図では、クライアントから送られてきたファイル108の内の一つに対して、共通ストア103内に存在するかを探し出す手順を示している。第一ステップは、クライアントから送られてきたファイル108のファイルサイズ(図18の例では32バイト)のファイル情報をファイル早見表109内から見つけだす。前述の通り、ファイル早見表109はファイルサイズでソートされているため、二分探索法による探索が可能である。図18の例ではファイルサイズが32バイトのファイルは三つある。第二ステップは、クライアントから送られてきたファイル108のチェックサムと同じ値のチェックサムを持つファイル情報を第一ステップで見つけたファイルの中から探し出す。前述の通り、ファイル早見表109はチェックサムでもソートされているため、この探索処理においても二分探索法が使える。図18の例ではチェックサムが一致するものは一つだけだった。第三のステップでファイル本体の比較を行う。第二ステップで発見したファイルの本体(図18の例ではファイルID=1001のもの)とクライアントから送られてきたファイル108の本体を比較する。

【0049】このような手順を踏むことで、ファイル本体の比較は、ファイルサイズおよびチェックサムが一致するものに限定できるため処理量を低減できる。なお図18の例ではファイルサイズにチェックサムを併用したが、チェックサム算出に要する処理負担が気になる場合には、ファイルサイズのみとしても良い。

### 【0050】3. 第三の実施例

第三の実施例はサーバ10とクライアントA30間のプロトコル(通信手順)に関するものであり、第一の実施例、第二の実施例に付け加える形で実現できる。以下では、第二の実施例の構成に基づいて説明する。

【0051】第三の実施例は、バックアップの際に行われるクライアントA30からサーバ10へのファイル転送時のファイル転送量を削減することを目的としている。そ

の実現のためにサーバ10上にすでにあると考えられるファイルは転送しない仕組みを持たせる。クライアントA30側のファイルがサーバ10側にすでにあるかどうかは、ファイルのパス名、サイズ、タイムスタンプなどのファイルに付属する情報、あるいはファイルの中身のチェックサムやCRC(巡回冗長符号)などで判断する。なおリストアについては第二の実施例と同じ仕組みが使える。

【0052】図19は第三の実施例を実装するシステム構成例である。図19の構成は、第二の実施例で出てきた図12の構成に転送量削減折衝処理S106(前記CPU41が前記メモリ42上で実行するプログラムとして実現される)と転送量削減折衝処理C304(前記CPU41が前記メモリ42上で実行するプログラムとして実現される)を加えた形になっている。第三の実施例では、図19の構成を前提として説明を進める。

【0053】図20は第三の実施例のサーバ10-クライアントA30間のバックアップ時のプロトコルを示したものである。クライアントA30-サーバ10間でやりとりされる情報の形式は、コマンドID(通知の内容を示す識別コード)+オプション情報(通知内容によって異なる)となっている。以下、その手順を追う。

【0054】最初にクライアントA30からサーバ10に対してファイルバックアップ要求801を発行する。それを受けたサーバ10はクライアントA30に対してファイル情報問い合わせ802を発行する。ファイル情報問い合わせ802の中には、要求する情報種別8021が含まれ、前記同一性チェックのための項目(ファイルパス名、サイズ、タイムスタンプ、チェックサム、CRC等)をクライアント側に知らせる。たとえばファイルのパス名とサイズ、タイムスタンプを同一性チェックの基準とするなら、ファイル情報問い合わせ802の中でファイルのパス名、サイズ、タイムスタンプと指定する。なお、ファイル情報問い合わせ802の中には、ファイルの範囲指定も可能で、全部あるいは指定範囲のみの二通りの設定8022が可能。指定範囲のみとした場合には、ファイルのパス名がその後に続く。ファイル情報問い合わせ802を受けたクライアントA30は、自身のファイル装置301内のファイルについて、ファイル情報問い合わせ802内で指定された項目についてファイル装置301の中を調べ、ファイル情報送信803としてサーバ10に通知する。ファイル情報送信803の中にはファイル情報8031が含まれており、さらにファイル情報8031の中にはファイルパス名80311と要求された情報80312が含まれている。要求された情報80312は、ファイルパス名80311に関する、要求する情報種別8021の情報(サイズやタイムスタンプ、チェックサム、CRC等)を入れる。ファイル情報送信803を受けたサーバ10は、ファイル情報送信803の中に含まれるファイル情報を調べ、共通ストア103や、第二の実施例に示す構成の場合のオーナー対応表1052に同じ内容のファ

イルが存在しないかを検索する。そして、ファイル情報送信803に含まれるファイル情報のうち、共通ストア103およびオーナー対応表1052（第二の実施例の場合）にないファイルのファイルパス名をファイル送信要求804として、クライアントA30に通知する。ファイル送信要求804を受けたクライアントA30は、要求されたファイル本体をファイル送信805としてサーバ10に送信する。サーバ10はファイル送信805を受けて、その内容を共通ストア103に格納するとともにオーナー対応表1052の内容を更新する。最後にサーバ10はクライアントA30に対して終了通知806を送信することでバックアップ処理を終了した旨を通知して終了する。

【0055】なお要求する情報種別が固定あるいはクライアント側で確定している場合には、ファイルバックアップ要求801、ファイル情報問い合わせ802をスキップし、ファイル情報送信803から始めても良い。また図20の場合、ファイル送信805がバックアップ処理の最後だと分かっているため、終了通知806も省略可能である。

【0056】前述の同一性チェック項目の内容はユーザによって変更可能な形を取っても良い。図21はその設定画面例である。大きな選択項目として、ファイル本体の照合が必要913か不要914かがある。同一性チェックの際、ファイルの内容についてもチェックする必要がある場合には、ユーザは必要913を選択する。ファイル内容のチェックが不要な場合には、ユーザが不要914を選択する。ユーザが必要913を指定した場合には第二の実施例と同等な動作となる。ユーザが不要914を選択した場合、さらに同一性チェックの際にファイルのどの項目を評価するかをユーザに選択させる。この画面で設定した内容をサーバ10内で記憶しておき、前記プロトコル内のファイル情報問い合わせ802中で要求する情報種別8021で指定するようにすれば良い。

【0057】図22は、図20に示すプロトコルを実現するクライアント側の処理フローチャートである。最初にサーバ10に対してバックアップ要求801を発行し、サーバ10からの応答を待つ。サーバ10からの応答が来たらその内容を調べ、終了通知なら終了する。またサーバ10からの応答が、ファイル情報問い合わせ802なら自身のファイル装置301内を調べてファイル情報送信803をサーバ10に通知し、再びサーバ10からの応答待ちに入る。サーバ10からの応答がファイル送信要求804ならファイル送信要求804内で指定されたファイルの本体を自身のファイル装置301から取り出し、ファイル送信805としてサーバ10に送信した上で、再びサーバ10からの応答待ちに入る。

【0058】同じく図20の第三の実施例の前記プロトコルを実現するサーバ10側の処理フローチャートを図23および図24に示す。

【0059】クライアントA30からのバックアップ要求8

01を受信したら、まずファイル本体の照合が必要かどうかを判定する（7210）。これは第三の実施例の前記同一性チェック設定（図21）で設定された内容を受けて判断する。必要911な場合には、クライアントに全ファイルの送信要求を発行し（7213）、以下、第二の実施例と同様に、ファイルをクライアントA30から受け取り、受信したファイルのそれぞれについて、共通ストア103内に同一内容のファイルが無いかどうかを確認しオーナー対応表1052と共通ストア103の内容を更新し（7214）、復帰（処理を終了）する。

【0060】不要912の場合には、バックアップ要求801を出したクライアントA30に全ファイルのファイル情報問い合わせ802を発行する（7211）。このファイル情報問い合わせ802の内容は、クライアント側処理の高速化のため図21で設定した内容から、チェックコードを除いたものにする。チェックサムやCRCなどのチェックコード処理を行っても性能的に問題無い場合には、7216、7217、7218、7219の処理を7211の処理中に行ってもよい。

【0061】クライアントA30から送られてくるファイル情報送信803を受信する（7212）。続いて同一見込みリスト1061を作成する（7215）。同一見込みリスト1061は図25に示すようにファイルのパス名の一覧であり、その記憶には一時的なメモリ領域（変数領域）を使用すれば良い。図23の処理7212で受け取った内容と、共通ストア103とオーナー対応表1052の内容を比較し、共通するファイルのパス名を同一見込みリスト1061に格納する。

【0062】続いてチェックコードによる確認が必要かどうかを判定する（7216）。必要でないなら、図24の処理（7221）に進む。必要な場合には、図23の処理7217、7218、7219を実行する。すなわち、同一見込みリスト1061中の各ファイルのチェックコードをクライアント側に要求する（7217）。言い換えると、処理7216では図23の処理7211、7212で照会した結果、同一であると考えられるファイルのチェックコードをクライアントに要求していることになる。そしてクライアントからチェックコードを受信し（7218）、共通ストア内のファイルのチェックコードと比較し、チェックコードが同じでないファイルを同一見込みリスト1061から取り除き（7219）、図24の処理（7221）に進む。

【0063】図24では、最初に同一見込みリスト1061に入っていないファイルがあるかどうかを判定し（7221）、ない場合には復帰（処理を終了）する。ある場合には、同一見込みリストに入っていないファイルの送信要求をクライアントに発行し（7223）、以下、第二の実施例と同様に、ファイルをクライアントA30から受け取り（7224）、受信したファイルのそれぞれについて、共通ストア103内に同一内容のファイルが無いかどうかを確認しオーナー対応表1052と共通ストア103の内容を更



新し(7225)、復帰(処理を終了)する。続いて、同一見込みリスト1061が空かどうかを判断し(7226)、空なら復帰(処理終了)。そうでないなら、同一見込みリスト1061内の各ファイルについてオーナー対応表の内容を更新(処理対象となっているクライアント名を追加)(7227)し、復帰(処理終了)する。

【0064】上述のように、第三の実施例によれば、クライアントからサーバにバックアップファイルを送信する際に、ファイル属性情報(ファイルパス名・ファイルサイズ・タイムスタンプなど)やファイル内容のチェックコード(CRC・チェックサム等)をサーバ・クライアント間で事前に交換することで、サーバ上にすでにあるファイルを再度送らずにすませることができ、バックアップ時のサーバの処理負荷、クライアントの処理負荷、ネットワーク負荷を低減し、バックアップ処理時間も短縮できる。

#### 【0065】4. 第四、第五の実施例の概要

第四、第五の実施例はリストア時に発生するサーバ10からクライアントA30へのファイル転送の方法に関するもので、第一の実施例、第二の実施例、第三の実施例のリストア処理に追加あるいはリストア処理を置き換える形で適用できる。以下、第四、第五の実施例に共通する内容について説明する。

【0066】クライアントA30のバックアップファイルをサーバ10に保管する場合、同じクライアントA30のバックアップ処理を繰り返すと、第三の実施例あるいは差分バックアップ等の技術を用いることで、クライアントA30からサーバ10への毎回のファイル転送量は低減される。一方、リストアでは対象クライアントのリストア用のファイルをすべて送るため、ファイル転送量は大きくなる。一般的にバックアップは定期的に行われるのに対して、リストアの頻度は低い。

【0067】すなわち、バックアップでは比較的小容量のファイル転送が多く行われるのに対して、リストアでは大容量のファイル転送が低い頻度で発生する。第四、第五の実施例はこのファイル転送の性質の違いを利用し、バックアップ時とリストア時とでファイル転送手段を変えるものである。

#### 【0068】5. 第四の実施例：記憶メディアによるリストア

以下、第四の実施例を第二の実施例をベースに説明するが、第一の実施例、第三の実施例にも適用可能である。

【0069】第四の実施例ではリストア時に発生するサーバ10からクライアントA30へのファイルの受け渡しに記憶メディアを使う。そのためサーバ10、クライアントA30双方に記憶メディアの読み書き装置を設ける。

【0070】図26はサーバ10のハードウェア構成を示したもので、図2に示す第一の実施例のハードウェア構成に記憶メディア書き込み装置47を追加している。記憶メディア書き込み装置47は記憶メディア49に情報を書き

込む機能を有するもので読み書き兼用であってもよい。記憶メディア49はたとえばMO、CD-ROM(書き換え可能型)、DVD(書き換え可能型)などの可搬型の記憶メディアが好ましいが、必ずしも可搬型である必要はなく、サーバ10からクライアントA30にデータ(ファイル)を受け渡し可能な記憶手段ならば良い。

【0071】図27はクライアントA30のハードウェア構成を示したもので、サーバ10の場合と同様、第一の実施例で取り上げた図2に記憶メディア読み込み装置48を追加したのになっている。記憶メディア読みとり装置48は、少なくとも記憶メディア49に格納されている情報を読み出す機能を有している。書き込み機能を有していても差し支えない。

【0072】図28は記憶メディアによるリストア処理のためのシステム構成とデータの流れを示した図である。サーバ10内に、オーナー対応表1052、共通ストア103、クライアント用メディア作成処理1062、メディア書き込み処理1063を備える。オーナー対応表1052、共通ストア103は図12に記したものと同一である。リストアデータ生成処理1062(前記CPU41が前記メモリ42上で実行するプログラムとして実現される)はオーナー対応表1052、共通ストア103を元にクライアントに渡すデータを生成する機能を持つ。メディア書き込み処理1063(前記CPU41が前記メモリ42上で実行するプログラムとして実現される)はリストアデータ生成処理1062で作られたデータを記憶メディア49に書き込む機能を持つ。一方、クライアントA30内にはメディア読み取り処理3051とリストア処理3052(それぞれ前記メモリ42上に前記CPU41が実行するプログラムとして実現される)、ファイル装置301がある。メディア読み取り処理3051は記憶メディア49内に記憶されている情報を読み出す機能を持つ。

【0073】図29は記憶メディア49の記憶内容を示した図である。記憶メディア49内に複数台のクライアントの情報を含む場合と単独のクライアントの場合とでは構造が異なる。複数格納時491の場合には、オーナー対応表4911と共通ストアファイル4912を含んでいる。すなわち、記憶メディア49上でもクライアント間でのファイル共有手法が働くようにしている。なお代案として、複数のクライアントファイルを個別に格納する形(493)にしても良い。単独の場合には、リストア対象のクライアント用のファイルを入れる(492)。

【0074】図30はサーバ側の記憶メディア作成処理1062を示した図である。最初に全クライアントのリストアデータを記憶メディア49に格納するか、一部を格納するかをたとえば図31に示す選択画面中の全体915か指定916にてユーザに選択させる。指定916が選択された場合にはさらにどのクライアントかをユーザに選択させる。常にクライアント全体を選択する形で良いならこの選択画面は必要ない。次にユーザによって選択されたク

クライアントの台数が単数が複数かを判定する。複数ならオーナー対応表1052の中を調べ、指定された複数クライアントがオーナーとなっているレコードを抜き出し記憶メディア49に書き込み、さらにそのファイルの本体を共通ストア103から抜き出して記憶メディア49に書き込む。単数の場合には、オーナー対応表1052の中を調べ、指定されたクライアントがオーナーとなっているファイルの本体を共通ストア103から抜き出して記憶メディア49に書き込む。

【0075】図32はクライアント側のリストア処理3052の流れを示した図である。まず記憶メディア49に格納されている内容を調べ、オーナー対応表4911が含まれているかどうかを判定する。オーナー対応表4911が無いなら、単数のクライアントの情報が含まれていると判断し、特定クライアント用ファイル4921の内容をクライアントA30自身のファイル装置301にリストアして、リストア処理は終了となる。オーナー対応表4911があるなら、オーナー対応表4911内に登場するクライアントのうち、どのクライアントに対してリストア操作を行うかをたとえば図33のようなユーザへの確認画面を用いてユーザに確認する。またクライアントA30のOS (Operating System) がクライアント名を保持しているなどの理由で、クライアントA30自身がクライアント名を知る手段があるならその手段を使って得られたクライアント名を自動的に選択する形にしても良い。その場合にはユーザによるクライアント名の選択操作は不要になる。クライアント名が明らかになった所で、記憶メディア49内のオーナー対応表4911の内容を調べ、該当クライアント用のファイル(群)を見つけだし、共通ストアファイル4912から対応するファイルの本体を取り出し、ファイル装置301にリストアして終了する。

【0076】6. 第五の実施例：大容量通信回線利用によるリストア

第五の実施例はリストア時のサーバ10-クライアントA30間データ転送に、通信衛星利用回線のような大容量データ転送向きの通信回線を用いるものである。これは本発明でのリストアのような比較的大容量のデータ送信に向く。また、前記通信衛星経由の通信を含むマルチキャスト、ブロードキャストなどの多数の相手に同時にデータを転送することが可能な通信手段は、本発明のような複数のクライアントでデータを共有するような構造を持つデータの配布に向く。なお通信回線のデータ帯域に余裕があるならリストアデータ、すなわちオーナー対応表1052と共通ストア103を常時送信しておいて、クライアントはリストアの必要が生じた時に、送られているデータを受信するという形にしても良い。

【0077】第五の実施例ではデータ転送手段として通信衛星を用いるため、サーバ10側には通信衛星に対する送信装置、クライアントA30側には通信衛星からの受信装置を備える。サーバ10側のハード構成例を図34に、

クライアントA30側のハード構成例を図35に示す。

【0078】図36は通信衛星を用いたリストア部分のシステム構成例である。第四の実施例とほぼ同様な構成となっているが、第五の実施例では記憶メディア49の代わりに通信衛星61を用いる。それに合わせてサーバ10側にBS (Broadcast Satellite) 送信装置1065、クライアントA30側にBS受信装置3061を備える。処理内容は第四の実施例と同様である。

【0079】上述のように、第四、第五の実施例によれば、リストア処理時に発生するサーバからクライアントへのファイル転送手段として、MOディスクなどの記憶媒体または衛星通信などの大容量通信回線を用いることで、通常用いるサーバ・クライアント間のネットワークトラフィックを低減できる。

【0080】

【発明の効果】本発明によれば、複数のクライアント機が使用される環境において、クライアント機の記憶装置のバックアップデータサイズを縮小することができる。

【0081】さらにバックアップへの障害をなくすことでバックアップを促しディスククラッシュからの復旧に対応できるようになり、システムの管理コストを低減することが可能になる。

【図面の簡単な説明】

【図1】本発明によるクライアント向きバックアップシステムの原理図である。

【図2】本発明のハードウェア構成を示す図である。

【図3】本発明の論理構成図を示す図である。

【図4】バックアップのプロトコルを示す図である。

【図5】リストアのプロトコルを示す図である。

【図6】サーバ側のメイン処理を示す図である。

【図7】サーバ側のバックアップ処理を示す図である。

【図8】サーバ側のリストア処理を示す図である。

【図9】共有効果の確認画面例を示す図である。

【図10】クライアント別共有比率の表示画面例を示す図である。

【図11】共有状況表示画面例を示す図である。

【図12】第二の実施例の構成図を示す図である。

【図13】オーナー対応表の例1を示す図である。

【図14】オーナー対応表の例2を示す図である。

【図15】第二の実施例のバックアップ処理を示す図である。

【図16】第二の実施例の状況表示画面例を示す図である。

【図17】第二の実施例におけるファイル早見表の構成を示す図である。

【図18】第二の実施例におけるファイル探索手法の例を示す図である。

【図19】第三の実施例の構成図を示す図である。

【図20】第三の実施例のプロトコルを示す図である。

【図21】第三の実施例の設定画面例を示す図である。

【図 2 2】 第三の実施例のクライアント側処理を示す図である。

【図 2 3】 第三の実施例のバックアップ処理（前半）を示す図である。

【図 2 4】 第三の実施例のバックアップ処理（後半）を示す図である。

【図 2 5】 同一見込みリストの例を示す図である。

【図 2 6】 第四の実施例、サーバのハードウェア構成を示す図である。

【図 2 7】 第四の実施例、クライアントのハードウェア構成を示す図である。

【図 2 8】 第四の実施例、論理構成図を示す図である。

【図 2 9】 第四の実施例、記憶メディアの内容を示す図である。

【図 3 0】 第四の実施例、サーバ側記憶メディア作成処理を示す図である。

【図 3 1】 第四の実施例、メディア作成種別指定画面例

を示す図である。

【図 3 2】 第四の実施例、リストア処理の内容を示す図である。

【図 3 3】 第四の実施例、クライアント名選択画面例を示す図である。

【図 3 4】 第五の実施例、サーバのハードウェア構成を示す図である。

【図 3 5】 第五の実施例、クライアントのハードウェア構成を示す図である。

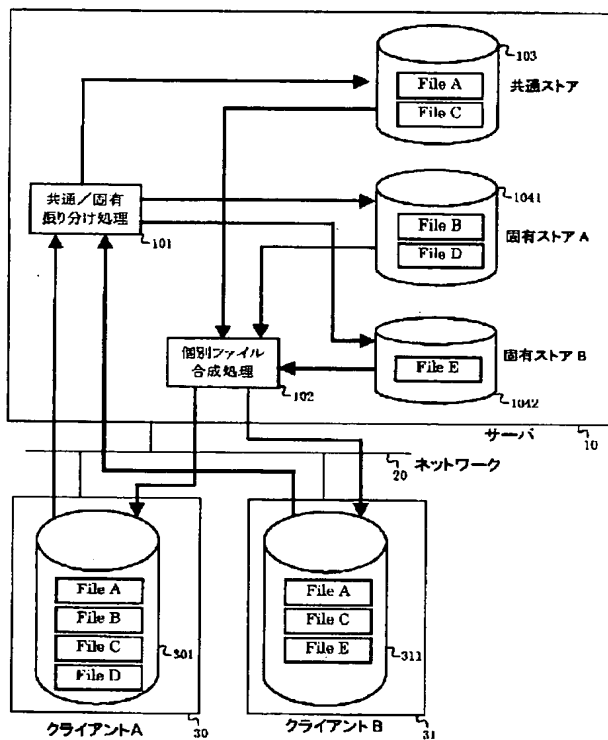
【図 3 6】 第五の実施例、通信衛星経由のリストアファイル転送を示す図である。

【符号の説明】

10…サーバ、101…共通／固有振り分け処理、102…個別ファイル合成処理、103…共通ストア、1041…固有ストアA、1042…固有ストアB、20…ネットワーク、30…クライアントA、301…クライアントA用ファイル装置、31…クライアントB、311…クライアントB用ファイル装置。

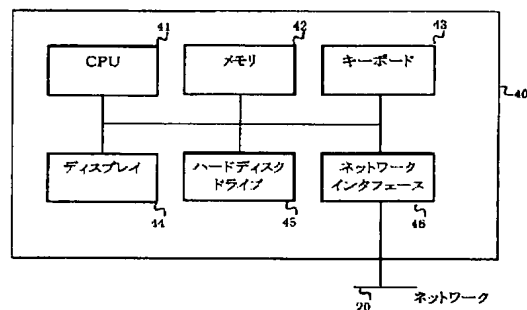
【図 1】

図 1.



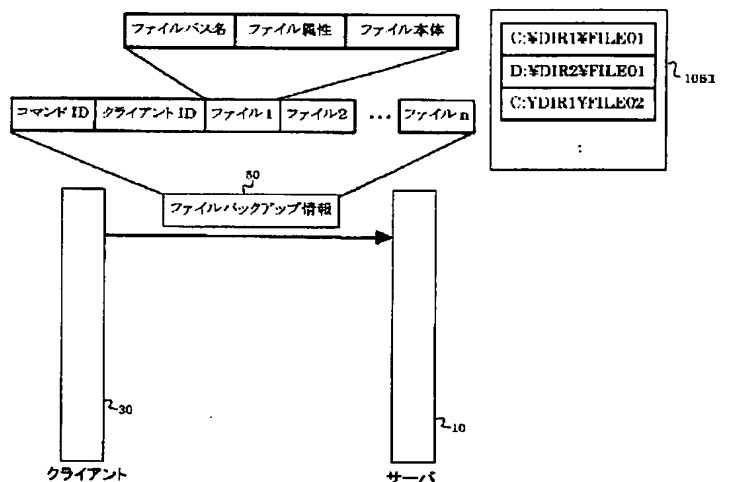
【図 2】

図 2.



【図 4】

図 4.

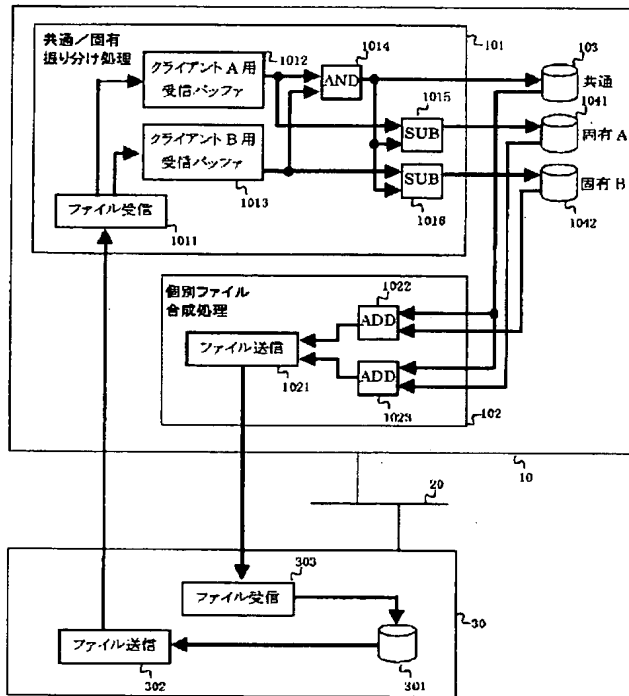


【図 2 5】

図 25.

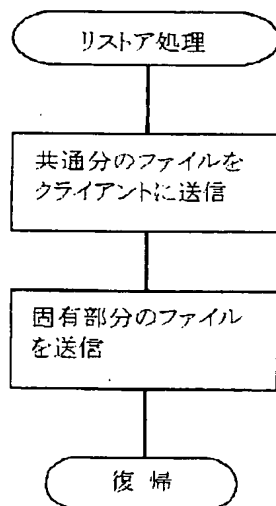
【図 3】

図3.



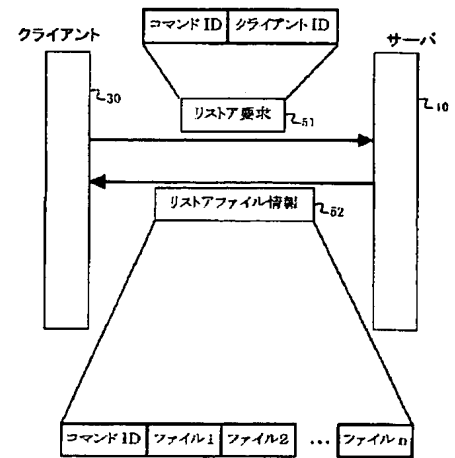
【図 8】

図8.



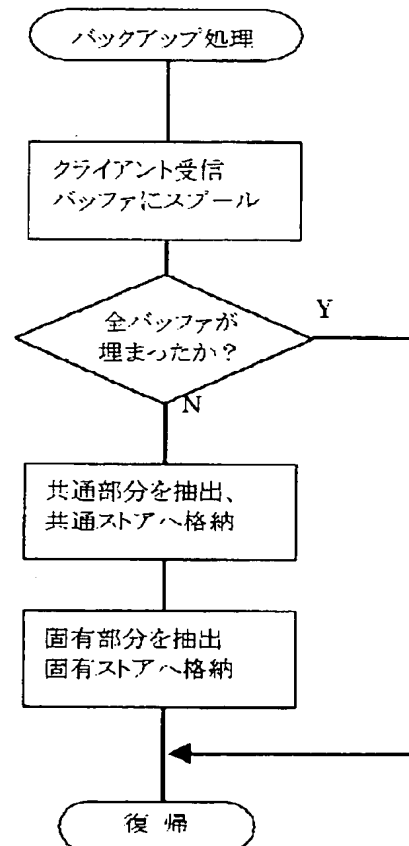
【図 5】

図5.



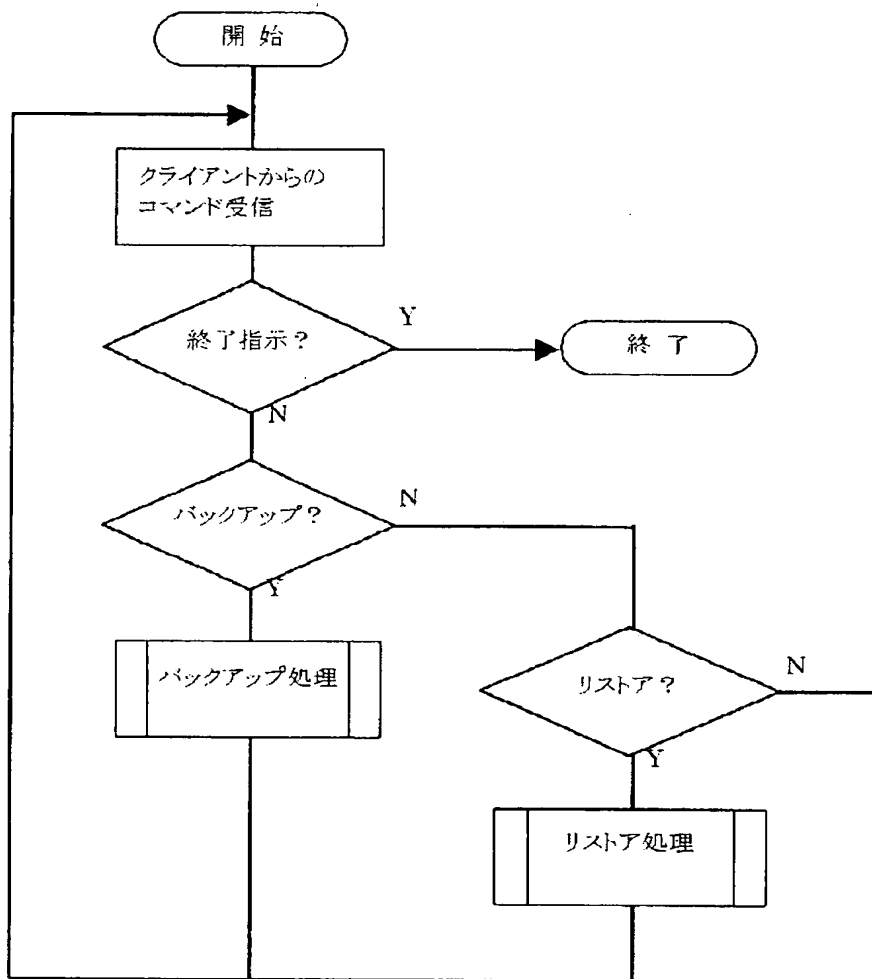
【図 7】

図7.



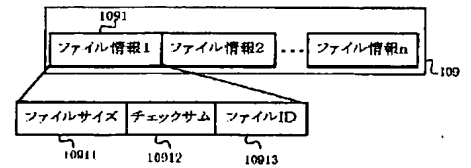
【図 6】

図 6.



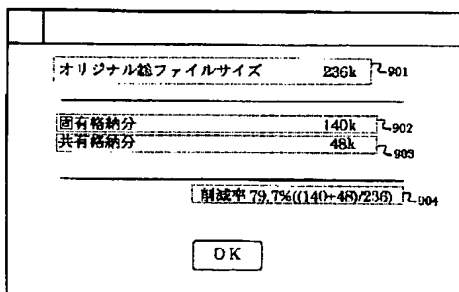
【図 17】

図 17.



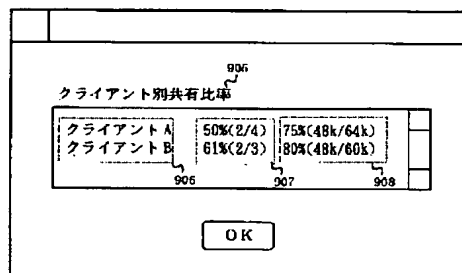
【図 9】

図 9.



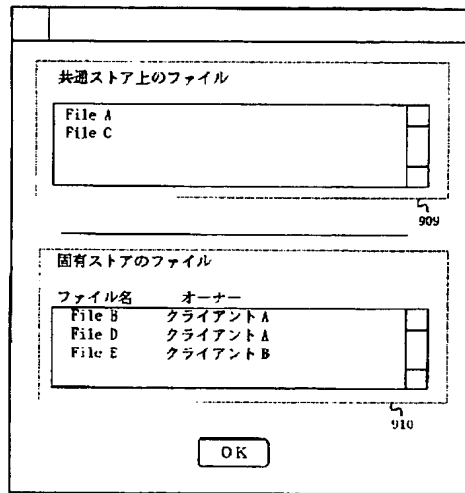
【図 10】

図 10.



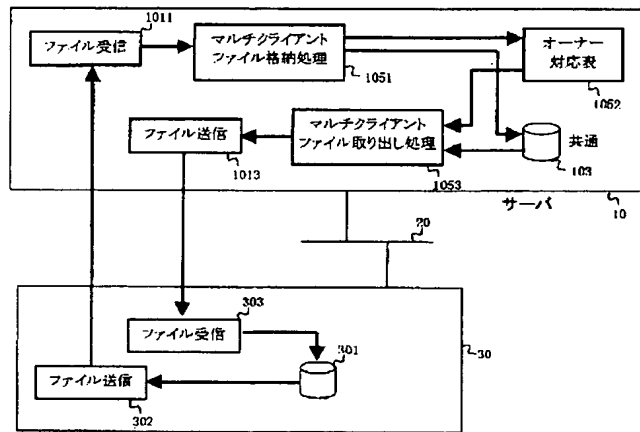
【図 11】

図 11.



【図 12】

図 12.

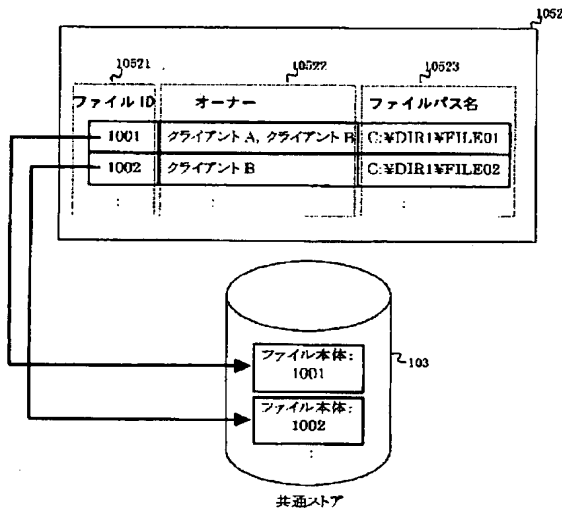


【図 14】

図 14.

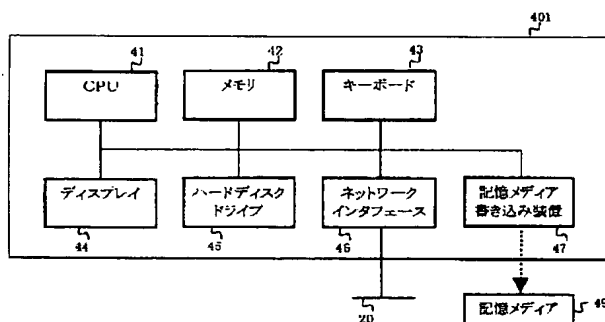
【図 13】

図 13.



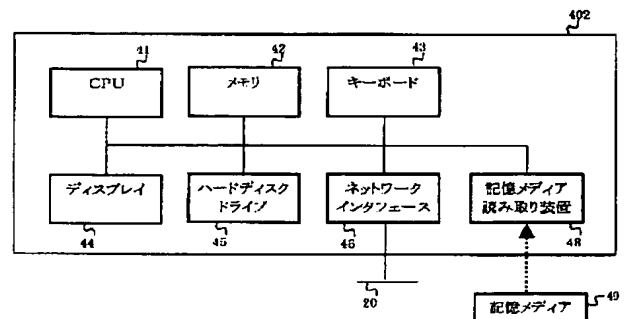
【図 26】

図 26.



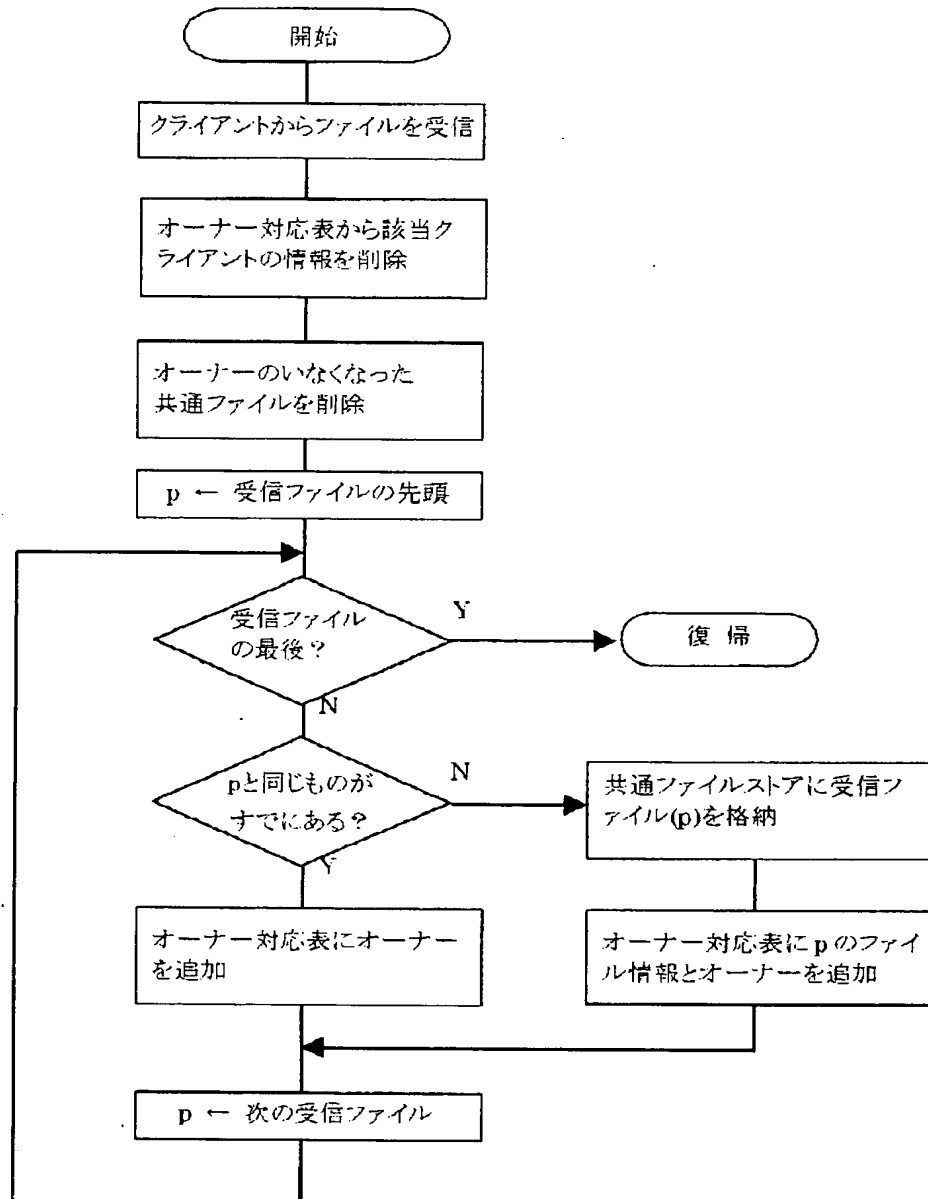
【図 27】

図 27.



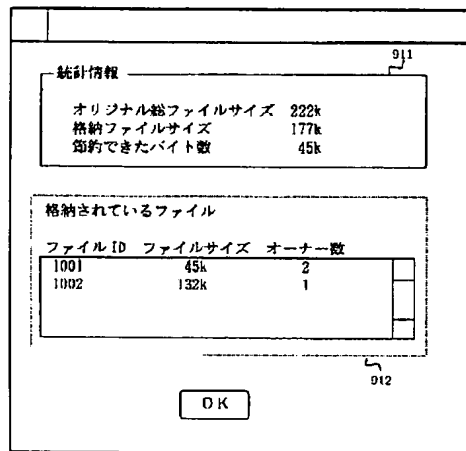
【図15】

図15.



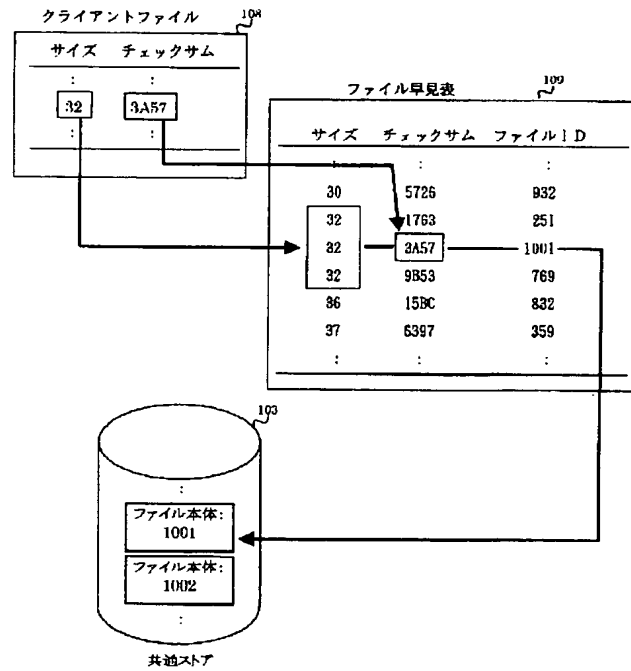
【図 16】

図 16.



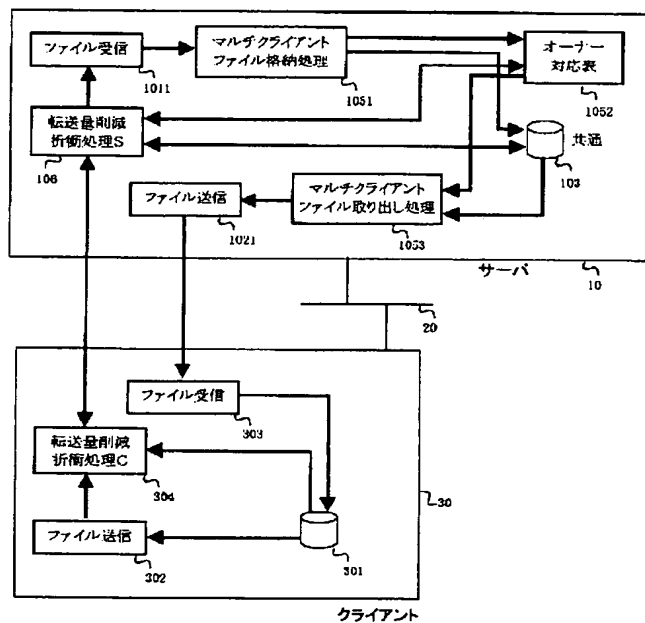
【図 18】

図 18.



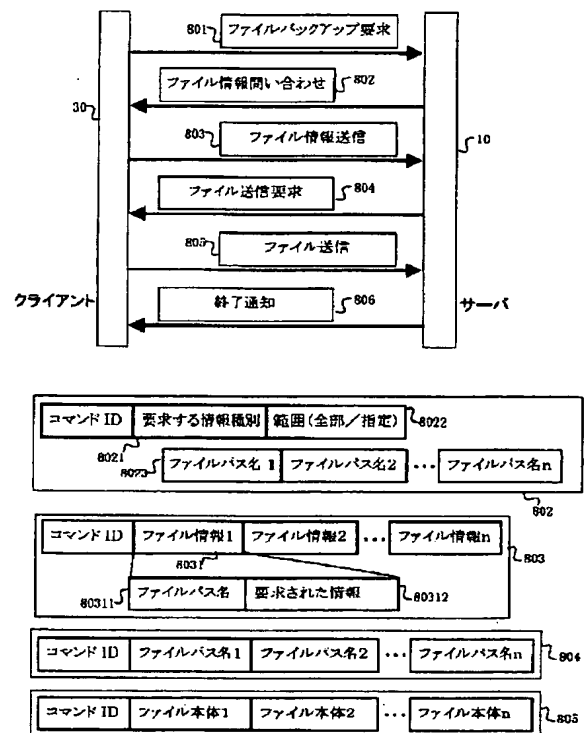
【図 19】

図 19.



【図 20】

図 20.





【図 21】

図21.

ファイルの照合内容の設定

ファイル本体の照合

914 ☐ 必要 913

☒ 不要

詳細項目

☒ ファイル名 ☒ ファイルサイズ

☐ ファイルパス名 ☐ チェックコード

☒ ファイル更新日時 ☐ CRC

☐ ファイル属性 ☐ チェックサム

OK キャンセル

【図 29】

図29.

複数格納時 491

オーナー 対応表	4911
共通ストア ファイル	4912

単体格納時 492

特定クライアント用 ファイル	4921
-------------------	------

複数格納時(代案) 493

クライアント A 用 ファイル	4931
クライアント B 用 ファイル	4931
...	
クライアント n 用 ファイル	

【図 33】

図33.

リカバリ

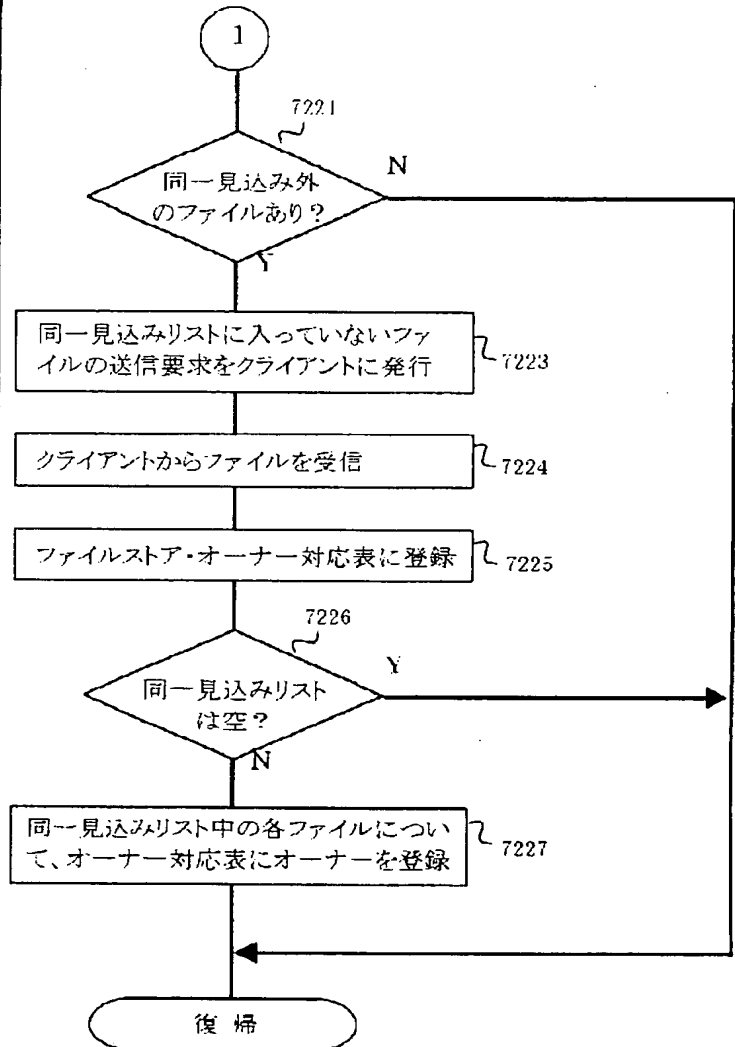
クライアント A

クライアント B

OK キャンセル

【図 24】

図24.



【図 31】

図31.

記憶メディアへの書き込み範囲の指定

☐ 全体 915

☒ 指定 916

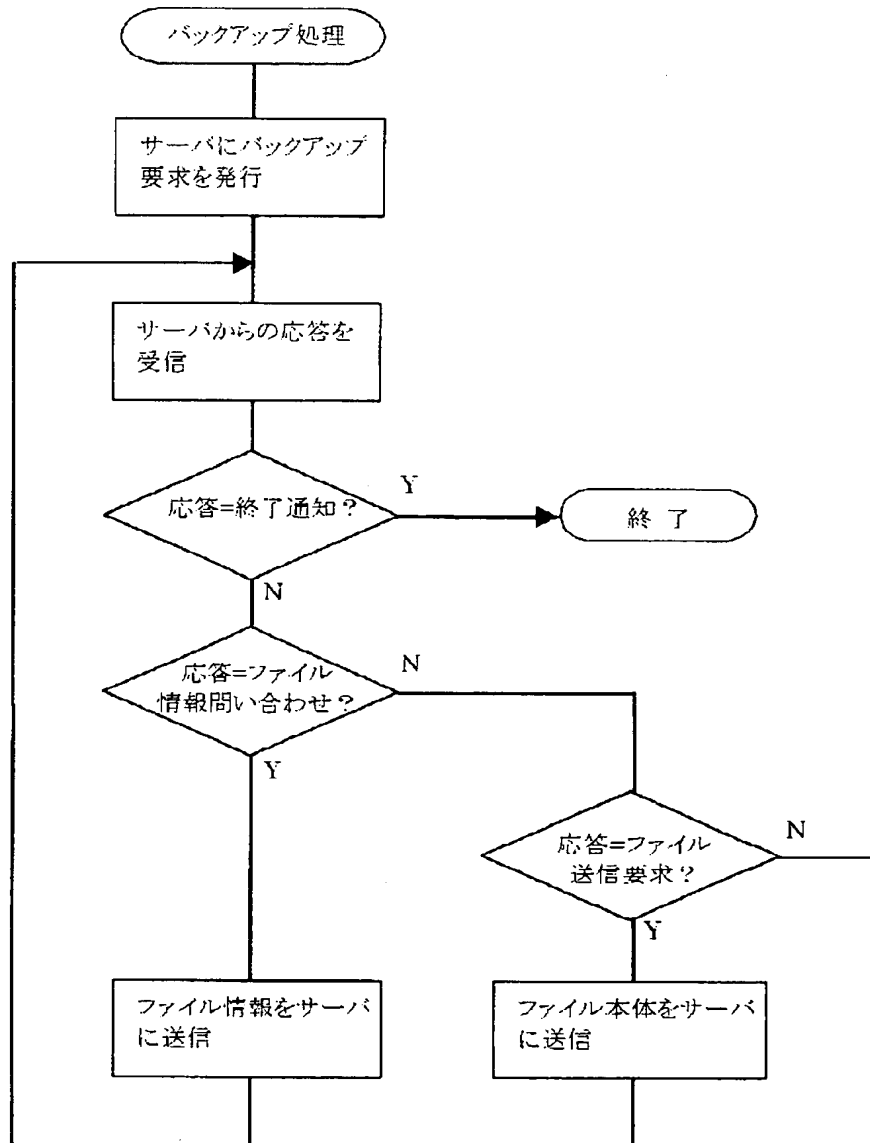
クライアント A

クライアント B

OK キャンセル

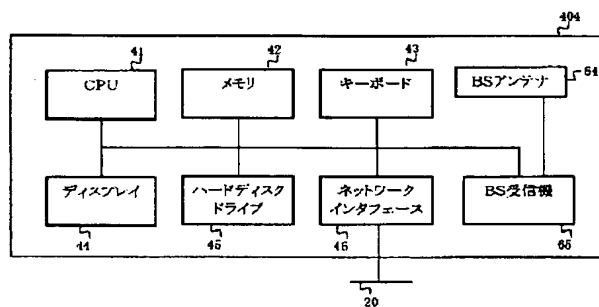
【図 22】

図 22.



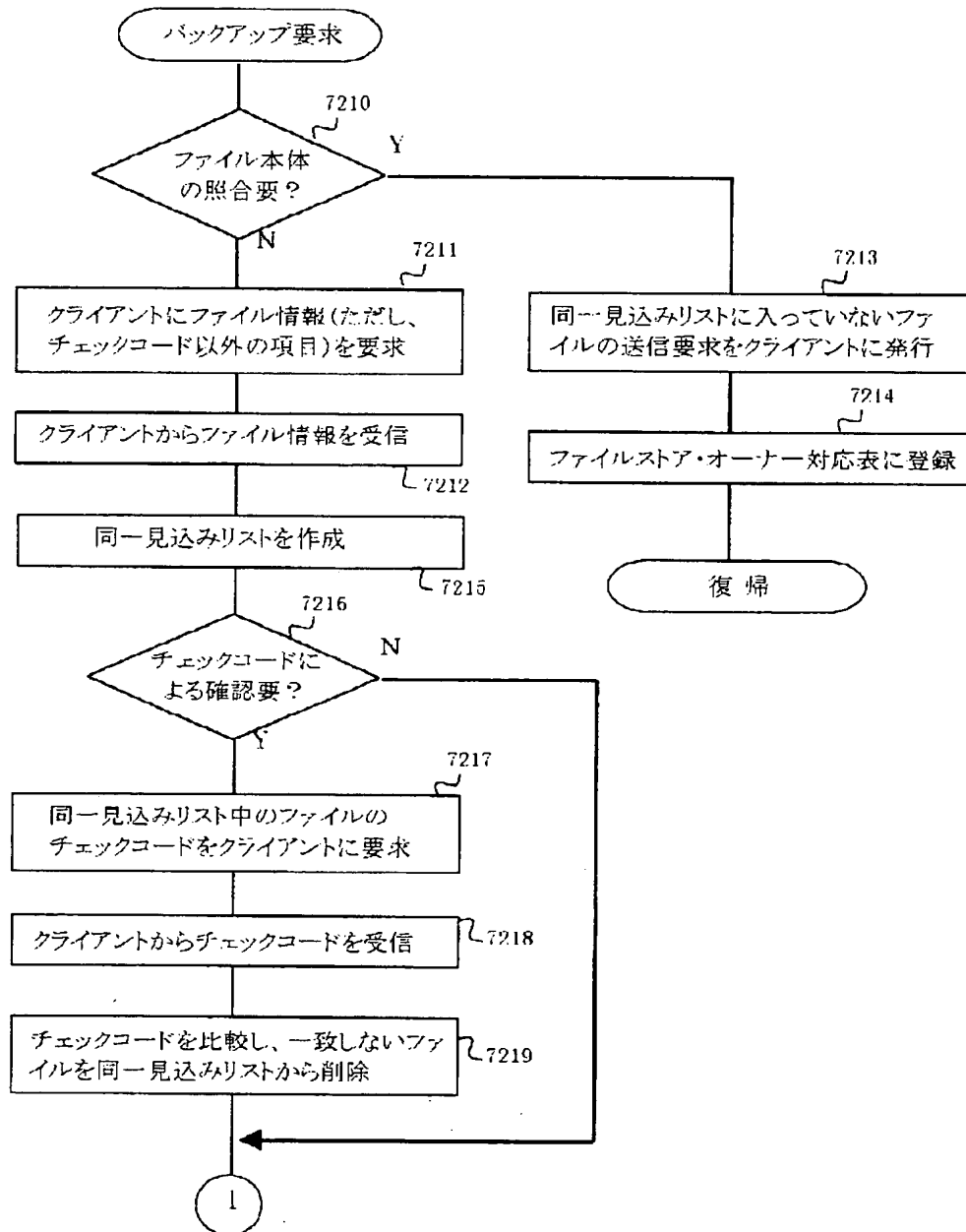
【図 35】

図 35.



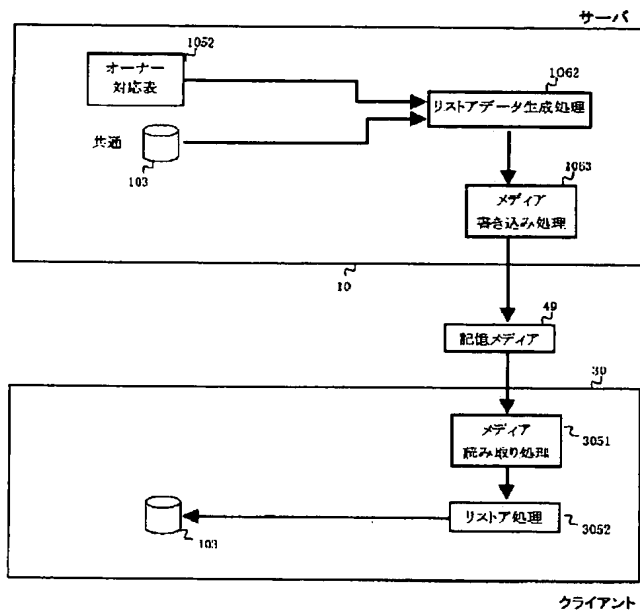
【図23】

図23.



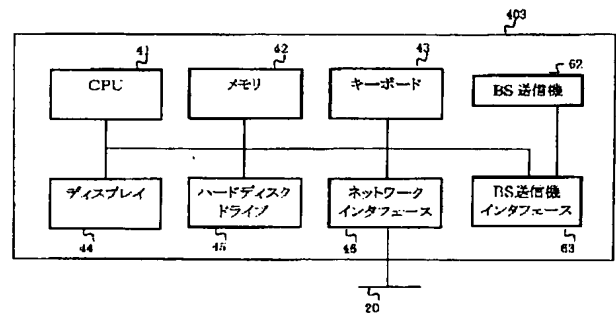
【図 28】

図28.



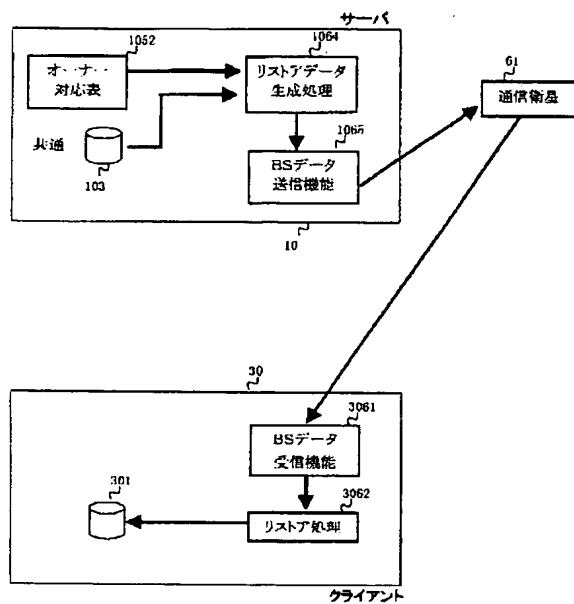
【図 34】

図34.



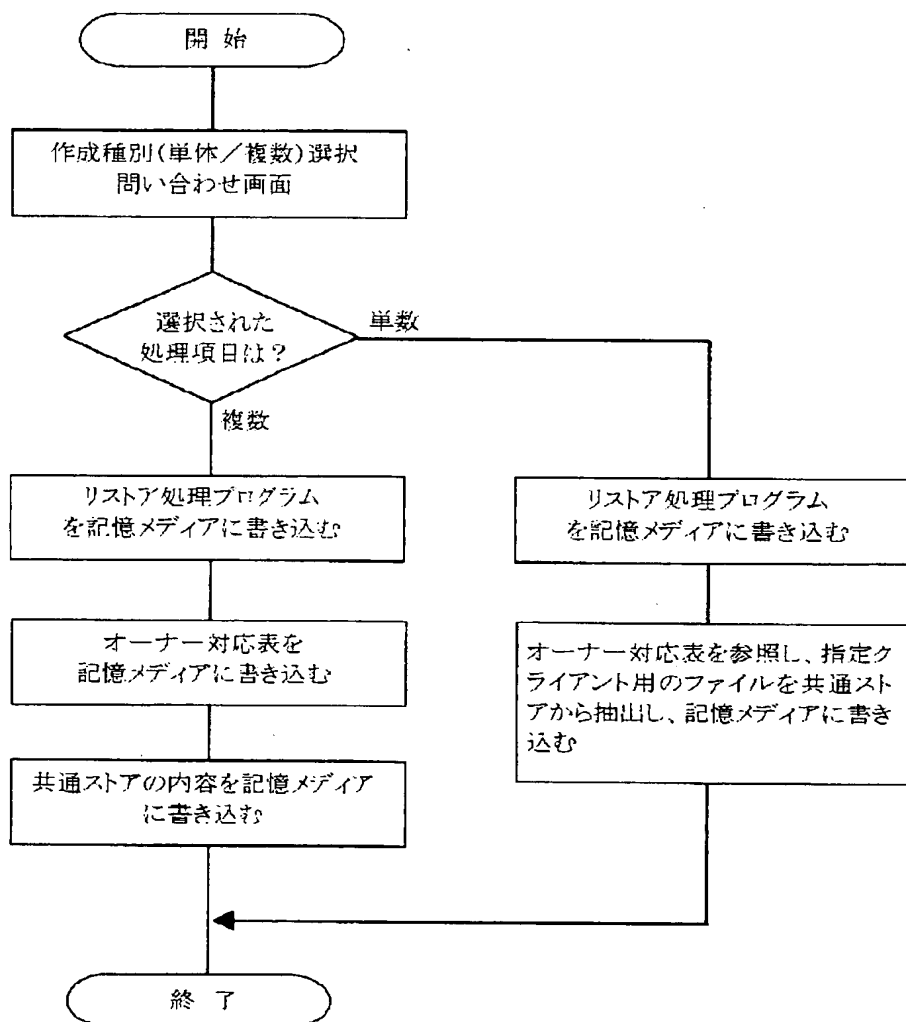
【図 36】

図36.



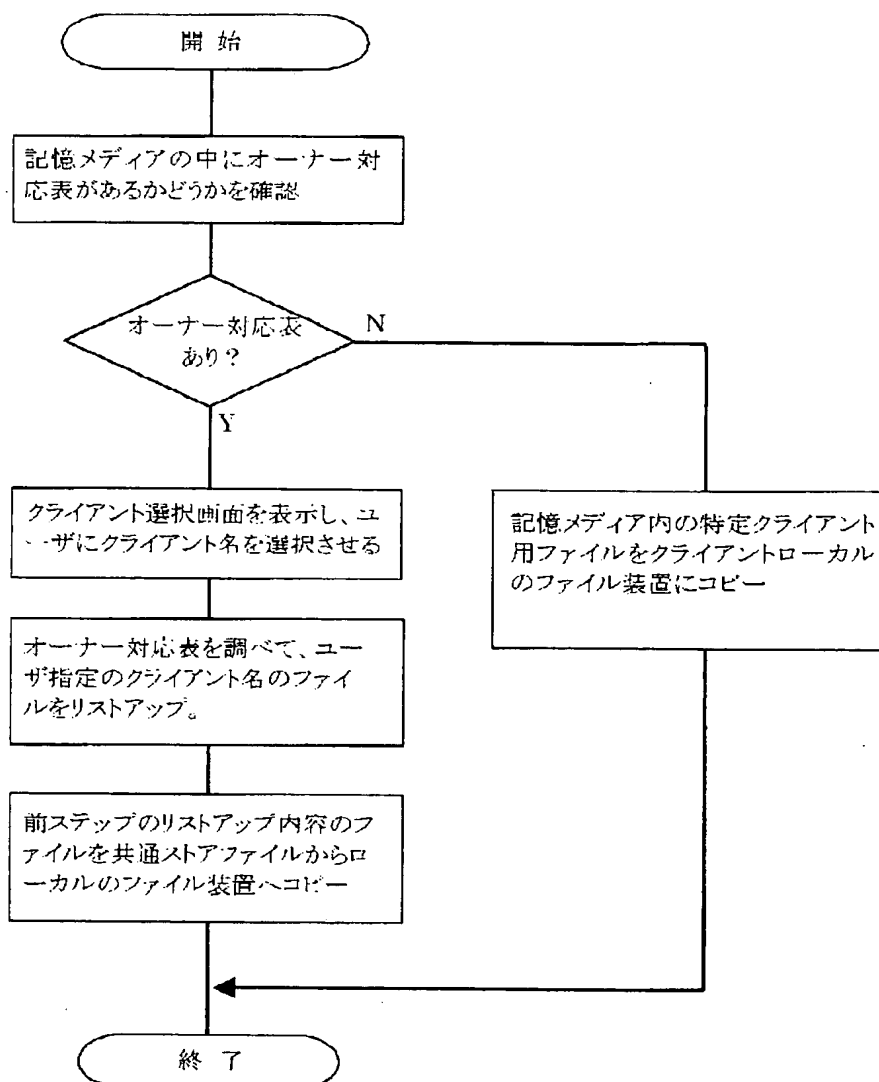
【図30】

図30.



【図32】

図32.



フロントページの続き

(72)発明者 齋藤 太朗  
神奈川県横浜市戸塚区戸塚町5030番地 株  
式会社日立製作所ソフトウェア事業部内

Fターム(参考) 5B082 DA02 DC05 DE06 HA08